ONLINE METHODS

All procedures conformed to local and US National Institutes of Health guidelines, including the US National Institutes of Health Guide for Care and Use of Laboratory Animals as well as regulations for the welfare of experimental animals issued by the German federal government.

Animals. Three male rhesus macaques were implanted with ultem headposts, trained via standard operant conditioning techniques to maintain fixation on a small spot for a juice reward and then scanned in a 3T Allegra (Siemens) horizontal bore magnet to identify face-selective regions using MION/Sinerem contrast agent (further details are provided in refs. 3,4). In all monkeys, a prominent face-selective region was located ~6-mm anterior to the inter-aural line. This middle face patch was targeted for recordings (details in ref. 4). Monkey A had a middle face patch located on the lip of the superior temporal sulcus, monkey T in the fundus and monkey L had two middle face patches, one on the lip (which we targeted) and one in the fundus.

Single-unit recording and eye-position monitoring. We recorded extracellularly with electropolished tungsten electrodes coated with vinyl lacquer (FHC). Extracellular signals were amplified, bandpass filtered (500 Hz to 2 kHz) and fed into a dual-window discriminator and an audio monitor (Grass). Spike trains were recorded at 1-ms resolution. Only well-isolated single units were studied. Cells that were visually responsive to the screening stimulus set (**Fig. 1b**) by ear were further tested with cartoon stimuli; in addition, some cells that were unresponsive to the screening stimuli by a formal criterion (see below) were also tested. Eye position was monitored with an infrared eye tracking system (ISCAN) at 60 Hz with an angular resolution of 0.25°, calibrated before and after each recording session by having the monkey fixate dots at the center and four corners of the monitor.

Visual stimuli. The monkey sat in a dark box with its head rigidly fixed and was given a juice reward for keeping fixation for 3 s in a 2.5° fixation box. Visual stimuli were presented using custom software (written in Microsoft Visual C/C++) and presented at a 60-Hz monitor refresh rate and 640 × 480 resolution on a BARCO ICD321 PLUS monitor. The monitor was positioned 53 cm in front of the monkey's eyes. Pictures subtended a 7° × 7° region of the visual field and cartoons subtended a 5.4° × 7.6° region on average, with both being presented at the center of the screen.

Pictures were presented for 200 ms, separated by 200-ms blank intervals in three experiments. In the first, 96 pictures from six different image categories (faces, human bodies, produce, technical objects, human hands and scrambled images) were shown (**Supplementary Fig. 1**). In the second, images of 16 real faces, 16 fitted cartoons, 16 technical objects and 16 cartoon face parts were shown (**Supplementary Fig. 1**). In the third, all 128 (2^7) decompositions of a cartoon stimulus were shown (**Supplementary Fig. 1**).

In contrast, cartoon stimuli were shown continuously, updated every seven frames (117 ms). Cartoon faces were defined by 19 parameters, each of which could take any of 11 values. The face defined by mean parameters ($p_1 = p_2 = \ldots = p_{19} = 0$) was specified by measurements taken from a photograph of Tom Cruise. Face aspect ratio defined the eccentricity of a solid ellipse constituting the face outline. Face direction defined the horizontal offset of the feature assembly (that is, eyes, eyebrows, nose and mouth) as a fraction of face width. Thus, the horizontal position of the feature assembly could range from the left edge of the face to the right. Height of feature assembly defined the vertical offset of the feature assembly as a fraction of face height. Hair was modeled as an inverted U of height, hair length and thickness, and hair width. Inter-eye distance was defined as the distance between iris centers, normalized by face width (ranging from almost cyclopean to abutting the edges of the face). Eye aspect ratio defined the aspect ratio and eye size the size of the ellipse surrounding the iris. Eye, iris and eyebrow were drawn only when the left (right) edge of the eye was to the right (left) of the left (right) edge of the face. Gaze direction defined 3 × 3 pupil positions in the eye as follows:

$$\begin{pmatrix} -4 & -3 & -2 \\ -1 & 0 & 1 \\ 2 & 3 & 4 \end{pmatrix},$$

where matrix position denotes iris position in the eye and matrix value denotes feature value. Parameter values 0, −5 and +5 all represented a straight gaze direction. Horizontal and vertical spacing between positions was fixed at 2 pixels. Iris size defined the size of a solid ellipse in the eye (with the same aspect ratio as the eye) as a fraction of eye size. The eyebrow was modeled as an angled line segment, with the angle defined by eyebrow slant, width defined by eyebrow width and height above eyes defined by eyebrow height, with the latter two being normalized by face width and face height, respectively. The nose was modeled as an outline of an isosceles triangle, with base width defined by nose base and altitude by nose altitude. The mouth was modeled as two half ellipses. In a smiling mouth, one half ellipse was black and the other was a gray mask, which served to carve out the curve of the upper lip; in a neutral/frowning mouth, both half ellipses were black and joined to form a convex mouth shape. The width of the mouth, expression of the mouth (smiling to frowning) and height of the mouth (open to closed) were defined by mouth size, mouth top and mouth bottom, respectively. The distance of the mouth below the nose was defined by the mouth-nose distance (this parameter only affected the vertical placement of the mouth; nose position was unaffected).

Picture data analysis. For each cell, we analyzed the poststimulus time histograms over 400 ms for all images shown (96 in first experiment, 64 in second and 128 in the third). Poststimulus time histograms were smoothed with a Gaussian kernel in time with $\sigma_t = 15$ ms. For experiments 1 and 2, we collapsed responses in each category to compute the cross-category response variance for each time bin. This variance had to be threefold higher than that of spontaneous activity (measured between σ_t and 80 ms after stimulus onset) for a cell to be classed as being visually responsive. The visual response period was then defined from the first to last point in time that exceeded the variance threshold. For cells that did not meet this strict criterion for visual responsiveness, a default response period was defined to last from 120 ms to 319 ms. Firing rates were computed as averages over this interval. In the case of the first experiment, the response magnitudes were determined for faces and objects relative to the baseline firing rate and normalized to the maximal response. A face selectivity index was then computed as the ratio between difference and sum of face- and object-related responses. For |face-selectivity index| > 1/3, that is, if the response to faces was at least twice (or at most half) that of nonface objects, a cell was classed as being face selective[45–47].

Face decomposition analysis. Because the 128 images in this experiment did not fall into distinct categories, the method for finding the response period deviated slightly from the procedure described above. The poststimulus response interval started when the firing rate exceeded a threshold equal to spontaneous activity plus two s.d. of spontaneous activity. The response interval ended when the response fell below this threshold value. The resulting 128-element response vector was subjected to a seven-way ANOVA with the presence/absence of each of the seven face parts (**Fig. 2a**) as factors.

Cartoon data analysis. All data analysis was performed using custom programs written in MATLAB (MathWorks).

Determining significance of tuning. For each cell and feature dimension, we computed time-resolved poststimulus tuning profiles (**Supplementary Fig. 2**) over three feature update cycles (351 ms of duration at 1-ms resolution) and 11 feature values. Profiles were subsequently smoothed with a two-dimensional Gaussian kernel of width $\sigma_t = 15$ ms in time and $\sigma_f = 1$ in the feature domain. We searched each profile for feature tuning, that is, increased diversity of response magnitudes, at each time delay. To minimize biases for tuning shape, we computed an entropy-related measure termed heterogeneity[48]. Heterogeneity is derived from the Shannon-Weaver diversity index $H' = -\sum_{i=1}^{k} p_i \log(p_i)$, with k being the number of bins in the distribution (11 in our case) and p_i being the relative number of entries in each bin. Homogeneity is defined as the ratio of H' and $H_{max} = \log(k)$; heterogeneity is defined as 1 – homogeneity. Thus, if all p_i values are identical, heterogeneity is 0, and if all values are zero except for one, heterogeneity is 1.

For each dimension and delay, we compared the heterogeneity value against a distribution of 5,016 surrogate heterogeneity values obtained from shift predictors. Shift predictors were generated by shifting the spike train relative to the stimulus sequence in multiples of the stimulus duration. This procedure preserved firing rate modulations by feature updates, but destroyed any

systematic relationship between feature values and spiking. From the surrogate heterogeneity distributions, we determined significance using Efron's[49,50] percentile method; for an actual heterogeneity value to be considered significant, we required it to exceed 99.9% (5,011) of the surrogate values. Note that this method is exact only for the actual 5,016 surrogate values and that a different set of values may have generated a different threshold. Therefore, to get a more robust and even more stringent significance level, we took as our significance threshold the average of the fifth largest heterogeneity value and the average of the five largest heterogeneity values. We validated this method with simulations, larger surrogate datasets of selected cells and by estimating significance levels from gamma functions fitted to the surrogate distributions (**Supplementary Fig. 9**). In the vast majority of cases reported here, the heterogeneity value of a significant tuning curve was much higher than even the largest of the surrogates. For a dimension to be considered significantly tuned, the significance threshold had to be passed at least twice at a temporal separation of at least $2\sigma_t$. We further required the tuning curve's maximal value to be at least 25% larger than the minimal value (see **Supplementary Text 5** and **Supplementary Figs. 10–12** for a different method, Gaussian fitting, for finding significantly tuned dimensions).

Co-occurrence of significant tuning. We found 14 out of 19 feature dimensions to be represented by middle face patch neurons. We then asked for each of the ${C_2}^{14} = 91$ feature combinations whether the frequency of occurrence was larger than would be expected by a model of chance associations. This model took into account the number of features each cell was tuned to (**Fig. 3c**) and the number of cells tuned to each feature dimension (**Fig. 3d**). We generated surrogate data that exactly matched these two distributions, but in which the associations between cell and features was otherwise random. Generating a distribution of 5,000 such surrogates for each feature combination, we tested for significance at $P = 0.0055$, a significance level at which only half a false positive dimension is expected on average.

Joint tuning functions were computed for a temporal delay suitable for both feature dimensions considered. For the analysis of interactions between significantly tuned dimensions, we first computed the center of mass of the heterogeneity measures (functions of time) of all significantly modulated dimensions to derive a 'joint delay'. When the optimal delays of both tuning curves considered were either shorter or longer than this joint delay, the center of mass between the heterogeneity functions of these two dimensions was chosen instead (**Supplementary Text 4** contains additional analysis of joint tuning functions).

Normalization conventions. For each cell, responses were baseline subtracted and divided by the maximal response above baseline; normalized responses were then averaged across cells. All tuning curves were normalized to the same area (**Figs. 4**, **6b** and **7c**, and **Supplementary Fig. 8**). The maximal response in the whole population of tuning curves was then set to 1 and the minimum was set to 0. In the inset in **Figure 4d**, deviating from this convention, the maximal response of each of the seven average tuning curves was set to 1 and the minimal response was set to 0.

46. Baylis, G.C., Rolls, E.T. & Leonard, C.M. Selectivity between faces in the responses of a population of neurons in the cortex in the superior temporal sulcus of the monkey. *Brain Res.* **342**, 91–102 (1985).
47. Perrett, D.I. *et al.* Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proc. R. Soc. Lond. B* **223**, 293–317 (1985).
48. Zar, J.H. *Biostatistical Analysis* (Prentice Hall, Upper Saddle River, New Jersey, 1998).
49. Efron, B. Bootstrap methods: another look at the jackknife. *Ann. Stat.* **7**, 1–26 (1979).
50. Manly, B.F.J. *Randomization, Bootstrap and Monte Carlo Methods in Biology* (CRC Press, Boca Raton, Florida, 2007).

doi:10.1038/nn.2363